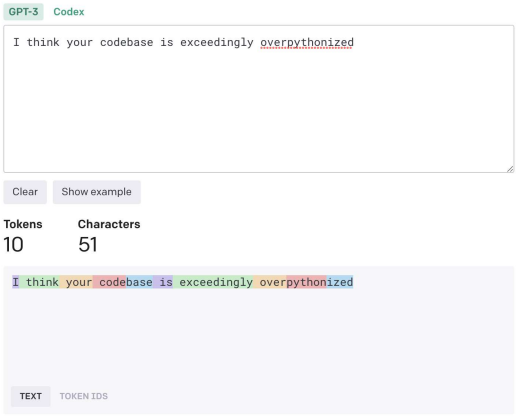
# AI-powered Social Media Post Generator

Creating the best fit social media post has always been a challenging task for many of us.It involves gathering data and knowledge and then rephrasing it From the steps like data and knowledge gathering to reframing it for the intended audiences.

GPT technology is popular for natural text generation. By harnessing these capabilities and fine-tuning them, we can significantly reduce the time and effort required for generating personalized post content.

GPT stands for Generative Pre-trained transformer - a neural network used for language processing. GPT technology uses Natural Language Processing (NLP) to turn data into human-readable text. NLP processes the data to generate text output, which is displayed using Natural Language Generation (NLG) and supported by Natural Language Understanding (NLU) for establishing relationships between input prompts to output. This process is essential during content generation testing.

- LLM models trained over a vast amount of dataset sourced from the internet, which is called pre-training. For example, the GPT-4 model has 1.8 trillion parameters and is trained on 1 petabyte datasets from web texts, books, articles etc.
- **Tokenization :** it is a type of text encoding. The basic idea is to assign short codes for symbols that are often used, which helps GPT models to perform better.
- **Entropy :** defines the minimal theoretically possible number of bits per symbol for a given data source. It is equivalent to the average number of bits per symbol.




$$\mathrm{H}(X) = -\sum_{i=1}^{n} \mathrm{P}(x_i) \log \mathrm{P}(x_i)$$

A symbol of frequency `p` will be encoded with approximately `-log(p)` bits. Multiply that by its frequency and we get a formula for each symbol's **contribution to entropy**: `-p * log(p)`. The above formula is the sum of all those contributions.

- **Embeddings :** Embeddings are numerical representations of concepts converted to number sequence, which make it easy for computers to understand relationships between texts. Basically, it is an information representation of the semantic meaning of a piece of text.
- **Transformer Architecture :** This mechanism helps the model to process the different tokens in a sequence, capturing dependencies and relationships between words. It employs self attention and multi-head attention mechanisms to process data efficiently.
- **Fine Tuning :** Fine Tuning involves updating the parameters of a pre-trained language model to generate more personalized content and to get better performance.
- **LLM Fine Tuning on Hugging face :** With auto train, we can easily fine tune large language models (llms) on our own data.
    - Casual Language Modeling
    - Masked Language Modeling ( In making )
- LLM finetuning accepts data in csv format. There are two models for LLM finetuning : *generic* and *chat*.
- **Datasets :** data obtained through social platforms by the profile link of the user after that it will be scrapped and tokenized so can be used for fine tuning of the models.
    - Platform-specific data
    - Diverse post content
    - User profile
    - Past post data
    - Emojis, hashtags, sentiment
- **Linkedin Profile API :** there are two ways to fetch the data of user's profile
    - Free for all, but permissions needed for profile data access
    - For approved developers on paid subscription to Linkedin's consumer solution platform.

    - **python :** linkedin_api library to obtain profile data by entering profile id of required user.